Le théorème de Goldbach

G. H. Hardy

Indications

Nous adressons nos très vifs remerciements au Professeur Hardy pour la bienveillance avec laquelle il a répondu à notre invitation de publier, dans le Mathematic Tidsskrift, la conférence qu'il y a donnée, donnant ainsi aux lecteurs de la revue une occasion tant attendue de se familiariser avec les nouvelles méthodes d'investigation profondes dont le Professeur Hardy, en collaboration avec le Professeur Littlewood, a enrichi la théorie des nombres, et par lesquelles il leur a été donné d'énoncer et de résoudre des problèmes difficiles qui, jusqu'alors, étaient considérés comme tout à fait inaccessibles.

1. Le théorème de Goldbach

Le célèbre problème dont je me propose de parler ce soir est probablement aussi difficile que n'importe lequel des problèmes non résolus des mathématiques, et ma conférence ne pourra pas être entièrement facile. Je devrai me contenter de vous donner une idée approximative de la nature et de l'histoire du problème, ainsi que des idées générales qui ont guidé les mathématiciens du passé et du présent dans leurs efforts pour trouver une solution.

Tout nombre pair est la somme de deux nombres premiers : tel est le théorème de Goldbach. Je peux peut-être commencer par l'observation triviale que, si "premier" a le sens qu'il a dans les mathématiques modernes, le théorème est manifestement faux. Il échoue pour 2, qui est premier, mais n'est pas la somme de deux nombres premiers. De nos jours, nous ne considérons pas 1 comme un nombre premier, car, si nous le faisions, la décomposition d'un nombre en facteurs premiers ne serait pas unique. Nous devons donc insérer les mots "supérieur à 2" dans l'énoncé du théorème.

Le théorème est énoncé dans la correspondance de Goldbach avec Euler en l'an 1742. Il semble que Descartes [1] l'ait anticipé. Il fut conjecturé indépendamment par Waring un peu plus tard, et c'est, je crois, dans les *Meditationes algebraicae* (1770) de Waring que la conjecture apparaît pour la première fois sous forme imprimée. Chacun de ces auteurs semble avoir observé que, si tout nombre pair (supérieur à 2) est la somme de deux nombres premiers, alors tout nombre (supérieur à 5) est la somme de trois nombres premiers. Vous verrez plus tard pourquoi cette remarque apparemment triviale devrait m'intéresser.

Les preuves numériques de la vérité du théorème sont accablantes. Il a été vérifié jusqu'à 1000 par Cantor (1894-1895), jusqu'à 2000 par Aubry (1896-1903), jusqu'à 10000 par Hausner (1896) ; et

Conférence à la Société Mathématique de Copenhague le 6 octobre 1921.

1. En matière d'histoire, je me contente de suivre l'ouvrage du professeur L. E. Dickson, *Histoire de la théorie des nombres* (Washington, 1919, vol. 1, p. 421 et suiv.). Dickson attribue cependant à Descartes l'affirmation selon laquelle "tout nombre pair est la somme de 1, 2 ou 3 nombres premiers", et ici, le mot "pair" devrait assurément être supprimé. *Mat. Tidsskr. B. lgan.*

d'autres données numériques, concernant des nombres spéciaux ou des nombres de formes spécifiées, ont été accumulées par Ripert (1903), Cunningham (1906), et Shah et Wilson (1919). Mais la plupart des calculs modernes ont été dirigés vers un but plus ambitieux, celui de déterminer ou de vérifier une formule asymptotique pour le nombre de décompositions en nombres premiers d'un nombre pair donné n.

Je désigne par $\nu(n)$ le nombre de façons dont $n = 2N$ peut être exprimé comme somme de deux nombres premiers, ou le nombre de solutions de l'équation

$$n = 2N = p + p'. \tag{1}$$

D'après le théorème de Goldbach, $\nu(n) > 0$ si $n \geq 4$; et un examen très rapide des preuves suffit pour rendre deux choses claires. En premier lieu, il y a beaucoup plus de vrai que ce qu'énonce le théorème. Non seulement $\nu(n)$ est positif, mais il est grand lorsque n est grand. Deuxièmement, bien que $\nu(n)$ tende vers l'infini avec n, il ne le fait pas de manière très régulière. La grandeur de $\nu(n)$ ne dépend pas seulement de la grandeur de n, mais aussi de sa forme arithmétique.

Il est important d'observer que ces conclusions sont en parfait accord avec le jugement *a priori* du sens commun. Nous raisonnons naturellement ainsi. Si $p < n$, et n est grand, la chance que p soit premier est approximativement $\dfrac{1}{\log n}$. Si alors nous écrivons n de toutes les manières possibles sous la forme $n = p + p'$, la chance que p et p' soient tous deux premiers est approximativement $\dfrac{1}{(\log n)^2}$. Nous devrions donc nous attendre à ce que l'ordre de grandeur de $\nu(n)$ soit

$$\frac{n}{(\log n)^2}. \tag{2}$$

D'autre part, la forme arithmétique de n est manifestement également pertinente. En premier lieu, il est évidemment pertinent de savoir si n est impair ou pair : si n est impair, il n'y a pas de représentation, à moins que $n = p + 2$, et alors il y en a une seule. Il n'est pas tout à fait aussi évident de considérer si n est un multiple de 3.

Nous devons cependant exclure tous les cas où soit p, soit p' est un multiple de 3. Si $3 \mid n$ [2], les deux ensembles de cas ainsi exclus sont les mêmes, tandis que si $3 \nmid n$, l'effet de la double exclusion est cumulatif. Nous devrions donc nous attendre à ce que la divisibilité par 3 augmente $\nu(n)$; et un raisonnement similaire s'applique à tout autre nombre premier, de sorte que $\nu(n)$ devrait être le plus grand lorsque n est composé d'un grand nombre de différents petits facteurs premiers. Toutes ces prévisions approximatives se révèlent être en parfait accord avec les faits.

Sylvester (1871) fut le premier mathématicien à suggérer une formule asymptotique pour $\nu(n)$. La règle de Sylvester est énoncée en mots, et une fois traduite en symboles, elle est la suivante :

$$\nu(n) \sim 2\pi(n) \prod_{p \mid n} \left(\frac{p-2}{p-1} \right), \tag{3}$$

2. Suivant Landau, j'écris $3 \mid n$ pour indiquer que n est divisible par 3, et $3 \nmid n$ pour exprimer le contraire.

où $\pi(n)$ est le nombre de nombres premiers ne dépassant pas n et le produit s'étend à tous les nombres premiers p pour lesquels $3 \leq p \leq n$ et $p \not| n$. Nous savons, contrairement à Sylvester, que

$$\pi(n) \sim \frac{n}{\log n}. \tag{4}$$

C'est le célèbre *Primzahlsatz* (théorème des nombres premiers), et nous pouvons l'utiliser pour simplifier (3). Nous avons également besoin d'un théorème de Mertens (1874), selon lequel

$$\prod_{p \leq n} \left(1 - \frac{1}{p}\right) \sim \frac{e^{-C}}{\log n}, \tag{5}$$

où C est la constante d'Euler. Ce théorème, qui est beaucoup moins profond que (4), montre que

$$\prod_{p \leq n} \left(\frac{p-2}{p-1}\right) = \prod_{p \leq n} \left(1 - \frac{1}{(p-1)^2}\right) \prod_{p \leq n} \left(1 - \frac{1}{p}\right) \sim \frac{2Ae^{-C}}{\log n},$$

où

$$A = \prod \left(1 - \frac{1}{(p-1)^2}\right), \tag{6}$$

le produit s'étendant maintenant à tous les nombres premiers impairs. Nous obtenons ainsi la formule

$$\nu(n) \sim \frac{4Ae^{-C}n}{(\log n)^2} \prod_{p|n} \left(\frac{p-1}{p-2}\right), \tag{7}$$

où p est un diviseur premier impair de n, et c'est ce que nous pouvons raisonnablement appeler la formule de Sylvester.

La formule de Sylvester est certainement erronée. Elle est correcte, apparemment, dans ses parties les plus manifestement intéressantes, dans son ordre de grandeur brut, et dans le facteur oscillant irrégulier qui dépend de la structure arithmétique de n. C'est le facteur constant $4Ae^{-C}$ qui ne peut pas être correct. Il est intéressant de s'arrêter un instant pour considérer comment un tel résultat négatif peut être établi.

Il n'y a aucun mystère en ce qui concerne la valeur moyenne de $\nu(n)$. Il est tout à fait facile de prouver [3] que

$$\nu(2) + \nu(4) + \cdots + \nu(n) \sim \frac{n}{2(\log n)^2}. \tag{8}$$

Si maintenant nous supposons une formule asymptotique pour $\nu(n)$, nous pouvons la tester par sa compatibilité avec (8). Nous pouvons supposer une formule du type de Sylvester, mais avec un facteur constant non spécifié B, et nous trouvons, comme test de compatibilité, que $B = 2A$. Ainsi, la seule formule possible de ce type est

$$\nu(n) = \frac{2An}{(\log n)^2} \prod_{p|n} \left(\frac{p-1}{p-2}\right). \tag{9}$$

3. Le théorème proprement dit est dû à Landau (1900).

La formule de Sylvester doit être erronée, sinon dans son principe, du moins par un multiplicateur constant $2e^{-C} = 1,123\ldots$ [4] (en réalité, il faudrait vérifier la valeur exacte). La formule (9), en revanche, semble presque certainement correcte.

La contribution de Sylvester au problème passa inaperçue pendant près de 50 ans. Je me tourne maintenant vers les auteurs, Merlin, Brun et Stäckel, qui l'ont attaqué récemment, et en particulier vers Brun. Les travaux de M. Littlewood et moi-même appartiennent à un cercle d'idées différent, et j'en parlerai plus tard.

Les écrits de ces trois mathématiciens ont une caractéristique commune frappante : ils n'utilisent que des méthodes élémentaires. Vous pouvez me demander ce qu'est une méthode "élémentaire", et je dois expliquer précisément ce que j'entends par cette expression. Je ne veux pas dire une méthode facile ou triviale ; une méthode élémentaire peut être désespérément ingénieuse et subtile. J'utilise le mot dans un sens défini et technique, et en cela je ne fais que suivre l'usage courant des arithméticiens. Par méthode élémentaire, j'entends une méthode qui ne fait pas usage de la notion de fonction analytique. Et la question que je souhaite vous poser est la suivante : est-il raisonnable, dans l'état actuel des connaissances mathématiques, d'espérer obtenir une preuve élémentaire du théorème de Goldbach ?

Si je réponds négativement à cette question, comme je dois et vais le faire, si je dis que je suis contraint de considérer tous ces efforts comme voués à l'échec, j'espère que vous ne me comprendrez pas de travers. Je ne peux pas croire que les méthodes de Merlin et Brun soient suffisamment puissantes ou suffisamment profondes pour conduire à une solution du problème. Mais je suis très loin de vouloir dire que je considère leur travail comme dénué d'intérêt et de valeur. Il y a beaucoup de choses dans les travaux de Brun en particulier qui me semblent très belles, et certains de ses théorèmes devraient, je pense, trouver leur place dans tous les livres sur la théorie des nombres.

Nous devons cependant tenir compte à la fois de l'histoire et de la structure logique de notre sujet. Revenons donc un instant à son théorème central, le "Primzahlsatz" ou "théorème des nombres premiers" exprimé par l'équation (4). Il semble évident que celui-ci doit être, en tout cas, un théorème plus facile que le théorème de Goldbach. Aucune preuve élémentaire n'est connue, et l'on peut se demander s'il est raisonnable d'en attendre une. Or, nous savons que le théorème est à peu près équivalent à un théorème sur une fonction analytique, le théorème selon lequel la fonction zêta de Riemann [5] n'a pas de zéros sur une certaine droite [6]. Une preuve d'un tel théorème, qui ne dépendrait pas fondamentalement des idées de la théorie des fonctions, me semble extrêmement improbable. Il est téméraire d'affirmer qu'un théorème mathématique ne peut pas être prouvé d'une manière particulière ; mais une chose semble tout à fait claire. Nous avons certaines vues sur la logique de la théorie ; nous pensons que certains théorèmes, comme on dit, sont profonds, et d'autres plus proches de la surface. Si quelqu'un produit une preuve élémentaire du théorème des nombres premiers, il montrera que ces vues sont erronées, que le sujet ne s'articule pas de la manière que nous avons supposée, et qu'il est temps de mettre de côté les livres et de réécrire la théorie.

4. Il est beaucoup plus difficile de vérifier l'écart par comparaison avec les faits. Shah et Wilson ont présenté une discussion très claire et intéressante de ce point.

5. $\zeta(s) = \zeta(\sigma + it)$.

6. $\sigma = 1$.

Vous êtes probablement familier avec l'idée générale du "crible" d'Ératosthène. Nous écrivons tous les entiers

$$\begin{gathered}\underline{1}, 2, 3, \underline{4}, 5, \underline{6}, 7, \underline{8}, \underline{9}, \underline{\underline{10}}, 11, \underline{12}, 13, \underline{\underline{14}}, \underline{\underline{15}}, \\ \underline{16}, 17, \underline{18}, 19, \underline{\underline{20}}, \underline{\underline{21}}, \underline{\underline{22}}, \ldots, x\end{gathered} \tag{8}$$

jusqu'à x. Nous effaçons (ou soulignons), d'abord 1 ; puis tous les nombres pairs après 2 ; puis les multiples de 3, sauf 3 lui-même ; et ainsi de suite, en répétant le processus pour chaque p. Le processus s'arrête lorsque p dépasse $\sqrt{x}$, car alors il ne reste que des nombres premiers.

Supposons maintenant que $p_1, p_2, \ldots, p_r$ sont les r premiers nombres premiers, que x est grand et r fixé, et que nous utilisons le crible pour ces seuls nombres premiers. Le nombre de nombres restants est approximativement

$$x\left(1-\frac{1}{p_1}\right)\left(1-\frac{1}{p_2}\right)\cdots\left(1-\frac{1}{p_r}\right);$$

il est facile de voir que l'erreur dans ce dénombrement est au plus $2^r + r$. Si $\pi_r(x)$ est le nombre de nombres ne dépassant pas x, et premiers ou non divisibles par aucun de ces r nombres premiers, alors

$$\pi_r(x) \sim x\prod_{i=1}^{r}\left(1-\frac{1}{p_i}\right).$$

Nous avons supposé jusqu'à présent que r est fixé. Le résultat serait beaucoup plus intéressant si nous pouvions supposer que r est une fonction de x. Supposons provisoirement que cela soit légitime, et prenons r comme le plus grand nombre premier ne dépassant pas $\sqrt{x}$. Nous obtenons

$$\pi(x) = \pi_r(x) \sim x\prod_{p\le\sqrt{x}}\left(1-\frac{1}{p}\right) \sim \frac{2e^{-C}x}{\log x}, \tag{10}$$

d'après la formule de Mertens (5). Cette formule est fausse (de sorte que notre supposition était illégitime), et il est significatif qu'elle soit erronée de la même manière que la formule de Sylvester (7). Notre échec peut contribuer à approfondir notre scepticisme quant aux résultats à attendre de l'utilisation du principe du crible.

Le "crible" utilisé par Merlin et Brun pour le théorème de Goldbach (le *crible de Merlin*) est d'un type légèrement plus complexe. Nous formons le tableau

$$\left.\begin{array}{cccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & \ldots & n-1 & (m), \\ n-1 & n-2 & n-3 & n-4 & n-5 & n-6 & n-7 & \ldots & 1 & (m'). \end{array}\right\} \tag{11}$$

Ici n est un nombre pair, et le tableau, lu verticalement, montre toutes les décompositions possibles $n = m + m'$ de n en entiers positifs. Nous effectuons maintenant le processus d'Ératosthène sur les deux lignes du tableau, en partant de la droite de la ligne inférieure, et en opérant avec les r premiers nombres premiers. Nous considérons qu'une décomposition $m+m'$ est effacée lorsque l'un de ses constituants est effacé.

Il y a deux cas à considérer pour chaque nombre premier p. Si $p \mid n$, les effacements dans la deuxième ligne tombent immédiatement en dessous des effacements correspondants dans la première, et le nombre de décompositions effacées est approximativement $\dfrac{n}{p}$. Si $p \not| n$, les effacements

ne correspondent jamais, et leur nombre est approximativement $\frac{2n}{p}$. Si alors $\nu_r(n)$ est le nombre de décompositions en nombres m, m' qui sont premiers ou non divisibles par aucun des r premiers nombres premiers, nous avons

$$\nu_r(n) \sim n \prod_{\substack{p|n \\ p\leq p_r}} \left(1-\frac{1}{p}\right) \prod_{\substack{p\nmid n \\ p\leq p_r}} \left(1-\frac{2}{p}\right). \tag{12}$$

Cette formule est correcte tant que r est fixé.

Supposons une fois de plus que cette formule est correcte lorsque p_r est environ $\sqrt{n}$. Nous obtenons

$$\nu(n) = \nu_r(n) \sim \frac{1}{2} n \prod_{p|n} \frac{p-1}{p-2} \prod_{p\leq\sqrt{n}} \left(1-\frac{2}{p}\right),$$

où la valeur $p=2$ est maintenant exclue. Mais, si $3 \leq p \leq \sqrt{n}$, nous avons

$$\prod_p \left(1-\frac{2}{p}\right) = \prod_p \left(1-\frac{1}{(p-1)^2}\right) \prod_p \left(1-\frac{1}{p}\right)^2 \sim \frac{16Ae^{-2C}}{(\log n)^3},$$

d'après la formule de Mertens (5). Nous obtenons ainsi

$$\nu(n) \sim \frac{8Ae^{-C}n}{(\log n)^3} \prod_{p|n} \left(\frac{p-1}{p-2}\right); \tag{13}$$

et c'est la formule à laquelle l'argument de Brun conduit naturellement [7]. La formule, comme celle de Sylvester, est erronée, et d'un facteur $4e^{-C} = 1,2\ldots$. On observera que la formule de Sylvester (7) est la moyenne géométrique entre (9) et (13).

J'ai expliqué que je ne crois pas qu'une preuve du théorème de Goldbach soit susceptible d'être trouvée par des méthodes telles que celles-ci. Mais il est certainement possible de prouver quelque chose de cette manière, et ce que Brun a prouvé me semble très intéressant en effet. Il a prouvé, par exemple, que tout grand nombre pair n peut être exprimé sous la forme $m+m'$, où m et m' sont des nombres composés d'au plus 9 facteurs premiers impairs, et que le nombre de ces décompositions est au moins de l'ordre de $\frac{n}{(\log n)^n}$. Sa méthode de preuve est suffisamment élémentaire, mais un peu compliquée, et l'idée qui la sous-tend peut probablement être expliquée plus clairement par référence à un problème plus simple.

Le nombre de nombres ne dépassant pas x et premiers ou non divisibles par aucun des nombres premiers $p_1, p_2, \ldots, p_r$ est (si $p_r < x$)

$$N = r + [x] - \sum \left[\frac{x}{p_\lambda}\right] + \sum \left[\frac{x}{p_\lambda p_\mu}\right] - \sum \left[\frac{x}{p_\lambda p_\mu p_\nu}\right] + \cdots, \tag{14}$$

7. Brun n'énonce pas explicitement la formule (13). Sa démarche consistait plutôt à développer son raisonnement jusqu'à obtenir une formule de ce type, puis à tenter de déterminer la constante par d'autres moyens. La détermination de l'unique constante possible par moyennage a été effectuée indépendamment par Stäckel, ainsi que par M. Littlewood et moi-même.

où $[x]$ est le plus grand entier contenu dans x. Ceci peut être montré immédiatement par la méthode d'Ératosthène, et il s'ensuit directement que

$$N > x\left(1-\frac{1}{p_1}\right)\left(1-\frac{1}{p_2}\right)\cdots\left(1-\frac{1}{p_r}\right) - A2^r, \tag{15}$$

où A est une constante. Aucune conséquence très intéressante ne peut en être tirée ; mais Brun a montré que, si nous nous contentons de multiplier le premier terme de droite de (15) par une constante inférieure à 1, nous pouvons réduire considérablement l'ordre du second, en fonction de r. Plus précisément, il prouve que

$$N > 0,3x\left(1-\frac{1}{p_1}\right)\cdots\left(1-\frac{1}{p_r}\right) - p_r^s. \tag{16}$$

Supposons maintenant que p_r soit environ $x^{1/6}$. Alors le premier terme est (d'après le théorème de Mertens) de l'ordre de $\dfrac{x}{\log x}$, tandis que le second est de l'ordre de $x^{5/6}$. Ainsi, N est au moins de l'ordre de $\dfrac{x}{\log x}$. Si un nombre ne dépasse pas x et que tous ses facteurs premiers sont supérieurs à $x^{1/6}$, le nombre de ses facteurs premiers est au plus 5. Nous sommes conduits au théorème selon lequel le nombre de nombres inférieurs à x et ayant au plus 9 facteurs premiers est au moins de l'ordre de $\dfrac{x}{\log x}$.

Le théorème est trivial, car Tchebychef a prouvé (et par des méthodes purement élémentaires) que le nombre de nombres premiers inférieurs à x est de cet ordre de grandeur. J'ai choisi ce théorème trivial, cependant, simplement comme une illustration simple, et Brun a prouvé beaucoup plus. En premier lieu, il a étendu l'argument aux nombres d'une progression arithmétique arbitraire $mk+l$ [8]. Ce théorème aussi n'est bien sûr pas nouveau, car nous savons, grâce aux travaux de de la Vallée Poussin, qu'une telle progression contient la quantité prescrite de nombres premiers. La méthode, en revanche, est dans ce cas très intéressante, car il n'y a pas de preuve élémentaire de l'un des théorèmes concernant les nombres premiers d'une progression arithmétique. Ce qui est encore plus important, c'est que Brun a pu traiter le "crible de Merlin" de la même manière, et en déduire le théorème très frappant et très beau que j'ai énoncé il y a un instant ; un théorème, puis-je ajouter, que M. Littlewood et moi-même sommes incapables de prouver analytiquement, avec l'équipement beaucoup plus puissant dont nous disposons.

Il semble clair cependant que, si nous voulons attaquer le problème principal avec une chance de succès, les méthodes de la théorie des fonctions sont indispensables ; et je vais maintenant essayer d'expliquer la méthode par laquelle M. Littlewood et moi l'avons attaqué. Les grandes lignes de l'argument sont assez évidentes et inévitables ; les difficultés résident dans sa mise en œuvre complète.

Nous écrivons

$$\Phi(x) = x^2 + x^3 + x^5 + \cdots = \sum_p x^p, \tag{17}$$

8. Le cas sans intérêt où k et l ont un facteur commun est naturellement exclu.

$$(\Phi(x))^r = \sum_n \nu_r(n) x^n, \tag{18}$$

de sorte que $\nu_r(n)$ est le nombre de représentations de n comme somme de r nombres premiers [9]. Nous avons

$$\nu_r(n) = \frac{1}{2\pi i} \int_C \frac{[\Phi(x)]^r}{x^{n+1}}\, dx, \tag{19}$$

d'après le théorème de Cauchy, le chemin d'intégration C étant le cercle $|x| = R$, où $0 < R < 1$. Nous prenons

$$R = 1 - \frac{1}{n}. \tag{20}$$

Nous ne basons pas notre analyse, cependant, sur les formules réelles que j'ai écrites. Il est plus commode d'utiliser les formules

$$f(x) = x^2 \log 2 + x^3 \log 3 + x^5 \log 5 + \cdots = \sum_p x^p \log p, \tag{21}$$

$$(f(x))^r = \sum_n N_r(n) x^n, \tag{22}$$

$$N_r(n) = \frac{1}{2\pi i} \int_C \frac{(f(x))^r}{x^{n+1}}\, dx. \tag{23}$$

Ici

$$N_r(n) = \sum \log p_1 \log p_2 \cdots \log p_r, \tag{24}$$

où la sommation s'applique à tous les ensembles $p_1, p_2, \ldots, p_r$ dont la somme est n. Le théorème de Goldbach affirme que $\nu_r(n)$ (ou, ce qui revient au même, que $N_r(n)$) est positif pour $n = 4, 6, \ldots$. Notre objectif est le plus complet de trouver une formule asymptotique pour $\nu_r(n)$; et il est facile de montrer que, si nous pouvons trouver une telle formule pour $N_r(n)$, nous pouvons en déduire une pour $\nu_r(n)$ par division par $(\log n)^r$ [10].

Notre idée fondamentale est la même que celle qui nous a guidés dans notre travail sur le problème de Waring. Le cercle unité est une barrière de singularités pour $f(x)$; nous pouvons dire, grossièrement, que $f(x)$ devient grand lorsque x approche du cercle. Il y a cependant certains points spéciaux du cercle au voisinage desquels $f(x)$ est le plus grand, et dont les contributions à l'intégrale (22) sont d'une importance dominante. Ces points sont les "points rationnels $\dfrac{h}{k}$", pour lesquels

$$x = e^{2\pi i h/k} = e\left(\frac{h}{k}\right) = x_{h,k}, \quad \frac{h}{k} = \frac{0}{1}, \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{3}{4}, \frac{1}{5}, \ldots$$

Il est évident, par exemple, que le point $x = 1$ doit être le plus important de tous ; car $f(x)$ est une série à coefficients positifs, et croît le plus rapidement lorsque x est positif. En général, on peut s'attendre à ce que l'importance du point $\dfrac{h}{k}$ diminue à mesure que k augmente ; et en fait

$$f(x) \sim \frac{\mu(k)}{\phi(k)} \left(\log \frac{1}{X}\right)^{-1}, \tag{25}$$

9. $v_r(n)$ n'a aucun lien avec le $v_r(n)$ de (12).
10. Si $0 < \delta < 1$, on a $(1 - \delta) \log n < \log p < \log n$ pour presque tous les nombres premiers considérés.

où $\mu(k)$ et $\phi(k)$ sont les fonctions arithmétiques bien connues de Möbius et d'Euler [11], si $x = Xe^{2\pi ih/k}$ et $X \to 1$ par valeurs positives.

Supposons qu'une formule du type de (24) ait été prouvée comme valable au voisinage de chaque point $\dfrac{h}{k}$. La procédure naturelle est alors la suivante. Nous divisons le cercle C en un grand nombre de petits arcs $\xi_{h,k}$, chacun associé à un point particulier $\dfrac{h}{k}$. Nous substituons à $f(x)$, dans la partie de l'intégrale (22) qui est prise le long de $\xi_{h,k}$, l'approximation dérivée de (24) ; nous évaluons les intégrales résultantes ; et nous obtenons ainsi une approximation pour $N_r(n)$ sous la forme d'une série infinie. Cette série, que nous appelons la série singulière, peut heureusement être sommée sous forme finie.

L'analyse à laquelle ce processus conduit est complexe et difficile. Il vaudra mieux que je commence par énoncer ce que nous pouvons prouver. En premier lieu, nous avons, pour des raisons qui apparaîtront plus tard, à supposer la vérité d'une hypothèse non prouvée. La célèbre hypothèse de Riemann peut être énoncée comme suit : la partie réelle d'un zéro de la fonction zêta ne dépasse pas $\dfrac{1}{2}$. Notre hypothèse est une extension naturelle de celle-ci. La théorie de la distribution des nombres premiers dépend, en général, de la théorie de $\zeta(s)$. Leur distribution dans une progression arithmétique $mk + l$, où l est premier à k, dépend d'un certain nombre de fonctions associées, désignées génériquement par $L(s)$. Ainsi, lorsque $k = 4$, il y a deux telles fonctions, définies respectivement par

$$L_1(s) = 1^{-s} + 3^{-s} + 5^{-s} + \cdots, \quad L_2(s) = 1^{-s} - 3^{-s} + 5^{-s} - \cdots. \tag{26}$$

La première est $(1 - 2^{-s})\zeta(s)$, mais la seconde est une nouvelle transcendante. Il y a $\phi(k)$ telles fonctions associées à un k donné.

La généralisation la plus naturelle de l'hypothèse de Riemann est

Hypothèse R. La partie réelle d'un zéro de $L(s)$ ne dépasse pas $\dfrac{1}{2}$.

Nous n'avons pas réellement besoin de toute la force de cette hypothèse, mais je me contenterai de l'énoncer sous sa forme la plus simple et la plus frappante.

Notre théorème principal est alors le suivant : Si l'hypothèse R est vraie, alors tout nombre impair suffisamment grand [12] est la somme de trois nombres premiers impairs. Le nombre $\nu_3(n)$ de représentations est donné asymptotiquement par la formule

$$\nu_3(n) \sim B\frac{n^2}{(\log n)^3} \prod_{p|n} \left\{ \frac{(p-1)(p-2)}{p^2 - 3p + 3} \right\}, \tag{27}$$

où

$$B = \prod_p \left(1 + \frac{1}{(p-1)^3} \right). \tag{28}$$

11. $\mu(k) = (-1)^m$ si k est le produit de m nombres premiers différents, $\mu(k) = 0$ sinon ; $\varphi(k)$ est le nombre de nombres inférieurs et premiers avec k.

12. C'est-à-dire, tout nombre de ce type à partir d'un certain point.

Le produit dans (26) s'étend aux diviseurs premiers impairs de n, et celui dans (27) à tous les nombres premiers impairs.

La preuve complète de ce théorème, et de théorèmes similaires concernant les représentations de nombres par quatre nombres premiers ou plus, paraîtra prochainement dans les *Acta Mathematica*. Pour l'instant, je vais essayer de vous expliquer seulement (1) comment la formule finale apparaît, (2) pourquoi il est nécessaire de supposer l'hypothèse R, (3) pourquoi notre méthode réussit pour trois nombres premiers ou plus, mais échoue pour deux.

Je ne prendrai pas ces questions dans l'ordre où je les ai énoncées : je commence par la seconde. Vous vous souviendrez qu'il est nécessaire d'approximer la fonction $f(x)$ au voisinage du point $\frac{h}{k}$. En prenant d'abord le cas le plus simple, supposons que $h = 0, k = 1, x_{0,1} = 1$.

Nous écrivons $x = e^{-y}$ et utilisons la formule bien connue de Mellin

$$e^{-y} = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} y^{-s} \Gamma(s)\, ds. \tag{29}$$

De là, nous déduisons

$$f(x) = \sum_p \log p\, e^{-py} = \frac{1}{2\pi i} \int y^{-s} \Gamma(s) \sum_p \frac{\log p}{p^s}\, ds, \tag{30}$$

qui est substantiellement (mais pas exactement)

$$-\frac{1}{2\pi i} \int y^{-s} \Gamma(s) \frac{\zeta'(s)}{\zeta(s)}\, ds. \tag{31}$$

Les intégrales sont prises le long d'une ligne parallèle à l'axe imaginaire et passant à droite du point $s = 1$.

L'intégrande a des pôles lorsque $s = 1$, lorsque s est nul ou un entier négatif, et à tous les zéros complexes de $\zeta(s)$, que nous désignons généralement par ρ. Si nous supposons l'hypothèse de Riemann, chaque ρ a une partie réelle $\frac{1}{2}$. Il est naturel de supposer que, si nous déplaçons le chemin d'intégration de plus en plus vers la gauche, et appliquons le théorème de Cauchy, nous exprimerons $f(x)$ sous la forme d'une série infinie, dans laquelle les termes les plus importants seront les termes correspondant aux résidus pour 1 et ρ. Si nous désignons ces termes par

$$\frac{1}{y} + \sum_\rho \frac{A_\rho}{y^\rho}, \tag{32}$$

alors le terme correspondant à ρ est d'ordre $y^{-1/2}$ lorsque y est petit, et nous pouvons espérer que l'ordre de la série sera à peu près le même. Dans ce cas, nous prouverons que $f(x) \sim y^{-1}$, qui est le cas le plus simple de (24), et que l'ordre de grandeur de l'erreur n'est pas notablement plus grand que celui de y^{-1}. Il est évident que notre estimation de l'erreur dépendra essentiellement de notre hypothèse.

Passant maintenant au cas général, je désigne par $\lambda(n)$ la fonction arithmétique de n qui est égale à $\log n$ lorsque n est premier et sinon à zéro. Nous avons maintenant

$$f(x) = f(Xe^{2\pi ih/k}) = \sum_n X^n \lambda(n) e\left(\frac{nh}{k}\right). \tag{33}$$

Nous écrivons

$$n = mk + l \quad (l = 1, 2, \dots, k;\ m = 0, 1, 2, \dots),$$

et nous obtenons

$$\left.\begin{aligned} f(x) &= \textstyle\sum_{l,m} X^{mk+l} \lambda(mk+l) e\left(\frac{lh}{k}\right) \\ &= \textstyle\sum_l X^l e\left(\frac{lh}{k}\right) \sum_m X^{mk} \lambda(mk+l). \end{aligned}\right\} \tag{34}$$

Ainsi, $f(x)$ est exprimé comme la somme d'un nombre fini de fonctions, dont chacune dépend manifestement de la distribution des nombres premiers dans une progression arithmétique $mk + l$. Nous sommes donc conduits à faire une hypothèse, analogue à l'hypothèse de Riemann, sur toutes les fonctions $L(s)$ dont dépend cette distribution ; en d'autres termes, nous sommes conduits à l'hypothèse R. À moins de faire une telle hypothèse, nos difficultés, déjà considérables, seront terriblement accrues.

Nous supposons donc l'hypothèse R. Nous écrivons maintenant $X = e^{-y}$ et exprimons $f(x)$, au moyen des formules (32) et (33), sous la forme d'une intégrale

$$\frac{1}{2\pi i} \int y^{-s} \Gamma(s) Z(s)\, ds, \tag{35}$$

où $Z(s)$ est une combinaison linéaire des dérivées logarithmiques $\dfrac{L'(s)}{L(s)}$ des fonctions L associées au module k. Le terme dominant de $f(x)$, au voisinage de $x_{h,k}$, est le résidu de l'intégrande pour $s = 1$, et un calcul simple montre que c'est la fonction qui apparaît dans le membre de droite de (24).

Nous revenons maintenant à l'intégrale (22), et considérons les contributions séparées des arcs $\xi_{h,k}$. Si nous substituons à $f(x)$, sur $\xi_{h,k}$, l'approximation donnée par (24), et transformons l'intégrale par les substitutions

$$x = Xe^{2\pi ih/k} = Xe\left(\frac{h}{k}\right), \quad X = e^{-y}, \tag{36}$$

nous obtenons

$$\left(\frac{\mu(k)}{\varphi(k)}\right)^r e\left(-\frac{nh}{k}\right) \frac{1}{2\pi i} \int y^{-r} e^{ny}\, dy. \tag{37}$$

Le chemin d'intégration est maintenant un segment d'une droite, parallèle à l'axe imaginaire dans le plan de y, et passant à droite de l'origine.

Nous pouvons substituer à ce segment la droite complète dont il fait partie, l'erreur introduite par cette approximation s'avérant sans importance. L'intégrale peut alors être évaluée en termes finis, et (36) prend la forme

$$\frac{n^{r-1}}{(r-1)!} \left(\frac{\mu(k)}{\varphi(k)}\right)^r e\left(-\frac{nh}{k}\right). \tag{38}$$

Nous devons finalement sommer par rapport à k et h, et nous obtenons la formule

$$N_r(n) \sim \frac{n^{r-1}}{(r-1)!} S, \tag{39}$$

où

$$S = \sum_{k,h} \left(\frac{\mu(k)}{\phi(k)}\right)^r e\left(-\frac{nh}{k}\right) = \sum_k \left(\frac{\mu(k)}{\phi(k)}\right)^r c_k(n), \tag{40}$$

$$c_k(n) = \sum_h e\left(-\frac{nh}{k}\right), \tag{41}$$

et h dans (40) est inférieur à k et premier à k.

La série S, la "série singulière", est la série dont dépend la solution de tous ces problèmes. Elle peut, comme je l'ai dit plus haut, être sommée sous forme finie [13]. Si r et n sont de parité opposée, alors $S = 0$. S'ils sont de même parité, alors

$$S = 2C_r \prod_{p|n} \left(\frac{(p-1)^r + (-1)^r (p-1)}{(p-1)^r - (-1)^r}\right), \tag{42}$$

où

$$C_r = \prod_p \left(1 - \frac{(-1)^r}{(p-1)^r}\right). \tag{43}$$

Si $r = 3$, nous sommes conduits à (26), et si $r = 2$ à (9) ; les puissances de $\log n$ qui apparaissent dans ces formules étant introduites dans le passage de N à ν, et il y a des formules similaires pour chaque valeur de r.

Il y a malheureusement une différence vitale entre le cas $r = 2$, qui correspond au théorème de Goldbach, et tous les autres. Nous devons remplir le squelette que je vous ai présenté, et le transformer en une preuve rigoureuse ; et ce faisant, nous nous trouvons contraints de supposer que $r > 2$. Il ne me reste plus qu'à vous expliquer brièvement la raison de cette limitation regrettable [14].

Notre travail repose sur un système de formules approchées pour $f(x)$, chacune valable près d'un point particulier du cercle unité. Il suffira pour mon propos de considérer le point $x = 1$. Si $x = e^{-y}$ et y est positif, notre formule pour $f(x)$ est de la forme

$$f(x) = \frac{1}{y} + O\left(\frac{1}{y^{1+\varepsilon}}\right),$$

où ε peut être n'importe quel nombre positif. Ainsi, $(f(x))^r$ consiste en deux parties, un terme y^{-r}, qui est exactement connu, et un terme d'erreur dont l'ordre est, d'après notre analyse, au moins $O(y^{-r/2})$; et chacun de ces termes donne naissance à un terme correspondant dans la formule finale. La contribution du terme dominant peut être trouvée précisément, et est d'ordre n^{r-1}. La

13. La sommation est un exercice simple mais amusant d'algèbre élémentaire, que je me contenterai de considérer comme allant de soi.

14. L'explication qui suit ne doit être considérée que comme une première approximation de la vérité.

contribution du terme d'erreur ne peut être estimée que de manière plus grossière, et le mieux que nous puissions en dire est qu'elle est d'ordre $O(n^{r/2-1})$. Cette erreur doit, si notre approximation doit réussir, être plus petite que le terme dominant ; et ceci exige que $r - 1 > \frac{r}{2} - 1$, ou $r > 2$. En fait, j'ai plutôt fait plus que justice à notre méthode. Il semblerait, d'après ce que j'ai dit, que nous soyons à la limite du succès. Mais il y a d'autres complications dans l'analyse, et nous échouons par une puissance de n.

Lorsque je discutais des méthodes de Merlin et Brun, j'ai laissé entendre qu'il n'était guère raisonnable de supposer qu'elles pourraient réussir. Vous pouvez naturellement me demander ce que je pense des perspectives des nôtres, et c'est une question à laquelle il m'est plutôt difficile de répondre. Vous devez me faire, pour l'instant, le cadeau de l'hypothèse R. Je présume que l'hypothèse de Riemann sera un jour prouvée. L'hypothèse R, j'en suis sûr, sera prouvée dans la semaine qui suivra, et les preuves seront substantiellement les mêmes. Rien ne suggère que, à cet égard, une fonction L se comporte différemment d'une autre.

En dehors de cela, je répondrais que, l'hypothèse une fois prouvée ou accordée, je ne vois aucune raison particulière pour que notre méthode ne réussisse pas. Elle me semble adéquate pour le problème ; les idées qui la sous-tendent ne sont pas trop faciles et sont suffisamment profondes. Elle échoue dans les détails, et non dans son principe, même en l'état ; l'échec n'est pas un échec de la méthode, mais des capacités analytiques de M. Littlewood et de moi-même. La méthode semble incarner les traits essentiels du problème, et conduit, naturellement et inévitablement, à ce qui est manifestement le résultat réel. Je crois que, lorsque le problème sera résolu, il le sera d'une manière telle que celle-ci.